

Measurement of WZ and ZZ production in pp collisions at 8 TeV in final states with b-tagged jets with the CMS experiment

C. Vernieri on behalf of the CMS collaboration
Scuola Normale Superiore, Piazza dei Cavalieri 7, Pisa
56127, Italy

Abstract

In this note we present a measurement of the VZ ($V=W,Z$) production cross section in proton-proton collisions at $\sqrt{s}=8$ TeV in the $VZ \rightarrow Vb\bar{b}$ decay mode with $V=Z \rightarrow (\nu\bar{\nu}, \ell\ell)$, $V=W \rightarrow \ell\nu$, ($\ell = e, \mu$). The results are based on data corresponding to an integrated luminosity of 18.9 fb^{-1} collected with the CMS experiment. The process is observed for the first time in this particular final state with a significance exceeding six standard deviations (σ). The measured cross sections are consistent with the predictions of NLO calculations.

1 Motivations

The study of VZ ($V=W,Z$) diboson production in proton-proton (pp) collisions provides an important test of the electroweak sector in the standard model (SM). The VZ production in pp collisions at the Large Hadron Collider (LHC) has been measured with CMS and ATLAS in fully and semi leptonic decay modes^{1,2}, while evidence for the $VZ \rightarrow Vb\bar{b}$ has been reported by Tevatron with 4.6σ significance³. Though the $b\bar{b}$ final state is not sensitive as the leptonic ones, it is interesting being the purest $b\bar{b}$ resonance which allows to test the b-jet identification and reconstruction. In addition the VZ final state is the the least reducible background for the $VH(b\bar{b})$ search, its measurement in the relevant phase space for the Higgs boson search represents a strong validation of the $VH(b\bar{b})$ analysis strategy⁴.

2 From VH to VZ

At LHC the most sensitive search for the Higgs boson decaying into a b quark pair is performed requiring the presence of an associated vector boson V decaying leptonically, to suppress the multi-jet QCD contribution and to provide an efficient trigger path (charged leptons, E_T^{miss}). According to the associated V decay mode, the corresponding analysis channels are the 0-lepton, 1-lepton and 2-leptons, where electron and muon channels are analyzed separately.

Given the $p_T(V)$ -dependency of the cross section, the $VH(b\bar{b})$ search requires a large boost (~ 100 GeV) for both the $b\bar{b}$ pair and the V boson to reduce the Z/W+jets and $t\bar{t}$ backgrounds. For each channel, several $p_T(V)$ bins with different signal and background contributions are analyzed separately in order to optimize the significance, see Table 1. The VZ measurement is then performed in the phase space defined by $p_T(V) > 100$ GeV, where approximately 15% of the WZ and 14% of the ZZ total inclusive cross sections are contained.

3 Analysis Strategy

The event selection is based on the reconstruction of the leptonic decay of the V and of the Z boson decay into two b-tagged jets. Dominant backgrounds to VZ production originate from $V+$ jets, $t\bar{t}$, single top, multi-jet (QCD) and VH production. Besides requiring requiring a large boost of the

$p_T(V)$ bin	low	intermediate	high
0-lep	[100,130]	[130,170]	> 170
1-lep	[100,130]	[130,180]	> 180
2-lep			> 100

Table 1: $p_T(V)$ category definition for each decay channel. Values are in [GeV].

vector boson and the $b\bar{b}$ pair⁵, also the back-to-back topology and minimal additional jet activity requirements are used to reduce $t\bar{t}$ and single top contributions, while $M(b\bar{b})$ helps to separate VZ from V+jets and VH.

The reconstruction of the $Z \rightarrow b\bar{b}$ decay is made by selecting the pair of jets in the event with the highest $p_T(jj)$. The two jets are tagged as b-jets using the Combined Secondary Vertex (CSV) algorithm⁶. $W \rightarrow \ell\nu$ decays are identified by requiring a single isolated lepton and additional E_T^{miss} . The identification of $Z \rightarrow \nu\nu$ decays requires the E_T^{miss} and no isolated leptons in the event. $Z \rightarrow \ell\ell$ candidates are reconstructed by combining isolated, oppositely charged pairs of electrons or muons and requiring the dilepton invariant mass to be within a Z mass window of 30 GeV width. Detailed event selection is reported in⁷.

The analysis strategy main features are:

1. Extraction of the normalization of the dominant backgrounds from the data
2. Improvement of the $M(b\bar{b})$ resolution with b-jet energy specific corrections (regression)
3. More efficient discrimination of the signal through the use of a multivariate analysis.

3.1 Background Estimate

Control regions kinematically close to the signal region are identified in data and used to correct the Monte Carlo yields estimated for the main background processes: V+jets (split into 0/1/2b content at generator level) and $t\bar{t}$ production. A set of simultaneous fits is performed to several distributions of discriminating variables in the control regions, separately in each channel, to obtain consistent scale factors (SF) by which the Monte Carlo yields are adjusted. These scale factors account not only for cross section discrepancies in the theoretical predictions, but also for potential residual differences in physics object selection in the boosted phase space. Good agreement between data and simulation is found after applying the fitted SF in several control regions for all modes as shown by the dijet invariant mass distribution for the combination of all five channels, in all $p_T(V)$ bins reported in Fig. 1, (i).

3.2 Specific corrections for b-jet energy

An optimal $b\bar{b}$ invariant mass resolution is important to separate VZ and VH processes. The $M(b\bar{b})$ resolution is already improved by selecting the boosted phase space. A b-jet specific correction is derived in this analysis in an attempt to recalibrate to the true b-quark energy, by applying multivariate regression techniques similar to those used by the CDF experiment⁸. The regression is essentially a multi-dimensional calibration to the particle level - including neutrinos - which exploits all the relevant properties of a b-quark jet. This procedure also addresses naturally the problem of semi-leptonic b decays (35%), which results in lower response with respect to the light quark/gluon induced jets. A specialized BDT⁹ is trained on simulated signal events and provides a correction factor that improves both the b-jet energy measurement and its resolution.

Inputs are chosen among variables that are correlated with the b-quark energy and well measured. They include detailed jet structure information about tracks and jet constituents which differs from light flavor quarks/gluons jets. Information from B-hadron decays on the reconstructed secondary vertices are used as well as soft lepton from semileptonic decay when available, providing an independent estimate of the b quark p_T . For the $Z(\ell\ell)$ channel is exploited also the information carried by the variables related to the E_T^{miss} vector. Since there is no real E_T^{miss} in the event it works as a kinematic constraints for the momentum equilibrium in the transverse plane.

The improvement on the $M(b\bar{b})$ resolution when the corrected jet energies are used is approximately 15%, impacting the analysis sensitivity by 10–20%, depending on the specific channel. After correction the resolution of the $Z \rightarrow b\bar{b}$ di-jet invariant mass is about 10% and it results in a better separation of the VZ/VH processes as reported in Fig. 2 (i).

The regression technique has been validated also looking at the p_T balance between the di-lepton and the di-jet systems

distribution in a $Z(\ell\ell)+b\bar{b}$ enriched data sample ($\ell = e, \mu$). Exactly two central jets are required and then we use CSV to select $b\bar{b}$ candidates. Candidate $Z(\ell\ell)$ decays are reconstructed requiring $75 < M(\ell\ell) < 105$ GeV. A data to simulation comparison is shown in Fig. 2 before regression (ii) and after regression (iii), reporting an improved resolution and scale. The mean of the distribution is centered at unity after the regression and the resolution, defined as the sigma divided by the mean, (sigma and mean are evaluated through a fit with a Gaussian function) is improved by 20%. Fig. 1 (ii) shows the $M(b\bar{b})$ distribution after the non resonant backgrounds subtraction. The VZ signal is clearly visible with a yield compatible with the SM expectation and a significance of 4.1σ .

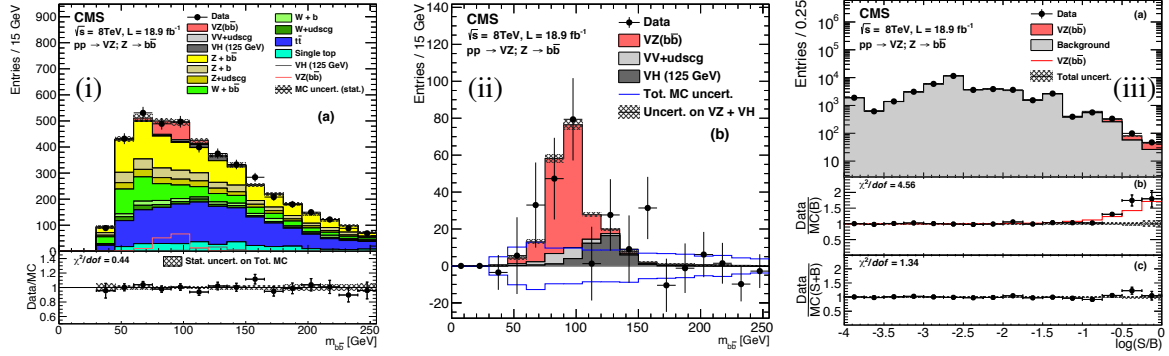


Figure 1: (i) Dijet invariant mass distribution. (ii) Same distribution after the non resonant subtraction. (iii) Events are sorted in bins of similar expected S/B as given by the BDT discriminant output value.

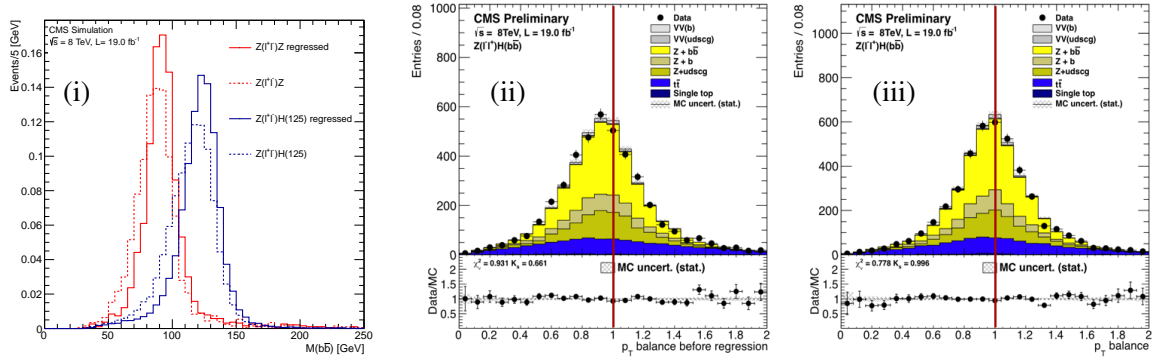


Figure 2: (i) Mass difference between ZH and ZZ simulated processed before and after the regression is applied. Distribution of the p_T balance on data versus MC before (ii) and after (iii) the regression.

3.3 Multivariate Analysis

A more efficient separation of the signal from the background is achieved by exploiting together with the improved $M(b\bar{b})$ other properties of the signal signature through a multivariate analysis. A BDT discriminant is trained using simulated samples for signal and all background processes. Among the most discriminating variables for all channels are the dijet invariant mass distribution, $M(b\bar{b})$, the number of additional jets, N_{aj} , the value of CSV for the jets and the distance in η - ϕ between the di-jet system.

The signal and background yields are then estimated in a fit region defined in the continuous output of the BDT, and a shape analysis is performed on that output.

Fig. 1 (c) combines all discriminants into a single distribution where all events, for all channels, are sorted in bins of similar expected signal-to-background ratio, as given by the value of the output of their corresponding BDT discriminants. The observed excess of events in the bins with the largest signal-to-background ratio is consistent with what is expected from VZ production. The VZ process

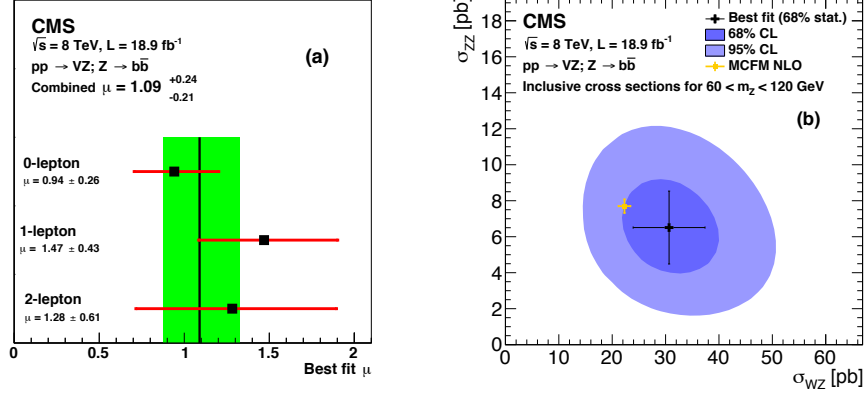


Figure 3: (a) The best-fit value of the VZ production cross section, relative to the SM cross section, for partial combination of channels and for all channels combined (band). (b) The 68%/95% CL contour regions for the WZ-ZZ production cross-sections.

is observed with a statistical significance of 6.3σ (5.9σ expected). This corresponds to a signal strength relative to the SM of $\mu = 1.09^{+0.24}_{-0.21}$.

4 Results

The total cross sections are determined by including all final states in a simultaneous constrained fit on the number of observed events in all categories. We extract the best fit value relative to the SM expectation at NLO. All cross section (relative to the SM value) extracted for individual channels provide compatible values with each other and the SM expectation Fig. 3 (a). To extract the WZ and ZZ cross-section a simultaneous fit floating both contributions independently is performed (Fig. 3-b). The best-fit is found to be at $\mu_{WZ} = 1.37^{+0.42}_{-0.37}$ and $\mu_{ZZ} = 0.85^{+0.34}_{-0.31}$. The resulting cross sections are measured to be:

$$\begin{aligned}\sigma(pp \rightarrow WZ) &= 30.7 \pm 9.3(\text{stat.}) \pm 7.1(\text{syst.}) \pm 4.1(\text{theo.}) \pm 1.0(\text{lumi.}) \text{ pb} \\ \sigma(pp \rightarrow ZZ) &= 6.5 \pm 1.7(\text{stat.}) \pm 1.0(\text{syst.}) \pm 0.9(\text{theo.}) \pm 0.2(\text{lumi.}) \text{ pb}\end{aligned}$$

for the Z boson produced in the mass region $60 < M_Z < 120$ GeV, the measurements are found in agreement with the NLO MCFM predictions.

References

1. CMS Collaboration, Physics Analysis Summary SMP-12-006;SMP-13-005;SMP-13-008 (2013).
2. Atlas Collaboration, Conference Note CONF-13-020, CONF-13-021 (2013).
3. CDF and D0 Collaborations, Phys. Rev. Lett. 109, 071804 (2012).
4. CMS Collaboration, Phys. Rev. D, 89, 1, pag 012003-012039, (2014).
5. J. M. Butterworth et al., Phys. Rev. Lett. **100**, 242001 (2008).
6. CMS Collaboration, JINST **8**, P04013 (2013).
7. CMS Collaboration, arXiv:1403.3047, submitted to *EPJC*, (2014).
8. T. Aaltonen et al., arXiv:1107.3026 (2011).
9. A. Hoecker et al., TMVA: Toolkit for Multivariate Data Analysis PoS A CAT 040 (2007).